

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicants: Bachar, et al.
Serial No.: 10/706,282
For: Apparatus and Method for Event-Driven Content Analysis
Filed: November 13, 2003
Examiner: Michael C. Colucci
Art Unit: 2626
Confirmation No.: 5641
Customer No.: 27,623

Attorney Docket No.: 0004794USU

AMENDMENT

Mail Stop Amendment
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In response to a non-final Office Action mailed April 26, 2010 ("the Office Action")
Applicants are submitting the present Amendment.

Amendments to the Claims begin on page 2.

Remarks begin on page 8.

IN THE CLAIMS

This listing of the claims should replace all prior versions:

1. (Currently Amended) An apparatus for event-driven content analysis of an audio interaction captured in a call center, within a computerized system having a processing unit and a storage unit, the apparatus comprising the elements of:
 - an audio or video recording device for recording the audio interaction and obtaining an interaction media;
 - a pivot spot defining component for automatically marking an at least one time position in the audio interaction that indicates the occurrence of an at least one pre-defined event or data item;
 - a first audio analysis component of a first audio analysis type;
 - a region of interest defining component for defining an initial region of interest, by determining the time limits of an at least one segment of the audio interaction, the segment containing the time position of a pivot spot, and for activating the first audio analysis component on the initial region of interest for dynamically reducing the time limits of the initial region of interest to obtain the region of interest; and
 - a second audio analysis component of a second audio analysis type for analyzing the region of interest of the audio interaction,
 - wherein the first audio analysis ~~component-type~~ and the second audio analysis ~~component-type~~ are selected such that the second audio analysis ~~component-type~~ requires more computing resources than the first audio analysis ~~component-type~~.
2. (Previously Presented) The apparatus of claim 1 further comprising a content analysis input selector component to determine an at least one input or parameter for the first audio analysis component or the second audio analysis component.

3. (Previously Presented) The apparatus of claim 1 further comprises an analysis type selector component to identify and to select the first audio analysis component or the second audio analysis component.
- 4-8. (Cancelled).
9. (Previously Presented) The apparatus of claim 1 wherein the first audio analysis component or the second audio analysis component is an audio analyzer component for analyzing audio elements of the interaction data.
10. (Previously Presented) The apparatus of claim 1 wherein the first audio analysis component or the second audio analysis component is a computer telephony interface events analyzer component for analyzing at least one computer telephony integration event occurring during the interaction.
11. (Cancelled).
12. (Previously Presented) The apparatus of claim 9 wherein the audio analyzer component further comprises the elements of:
 - a word spotting component to locate and identify pre-defined terms or patterns in the speech elements of the interaction data;
 - an emotion analysis component to locate and identify positive or negative emotions in the interaction data; and
 - a talk analyzer component to identify and locate specific pre-defined speech events in the speech elements of the information data.
- 13-14. (Cancelled).
15. (Previously Presented) The apparatus of claim 1 wherein the interaction media comprises at least one data packet carrying voice or other media over internet protocol.
16. (Previously Presented) The apparatus of claim 1 wherein the region of interest is a specific segment of the interaction media that is analyzed to extract meaningful interaction-specific information in an organization.

17. (Previously Presented) The apparatus of claim 1 wherein the interaction is associated with an at least one computer telephony integration event occurring during the interaction.
18. (Cancelled).
19. (Currently Amended) A method for event-driven content analysis, within a computerized system having a processing unit and a storage unit, the method comprising the steps of:
 - receiving an audio interaction media between an organization and a customer, the interaction media associated with an at least one event, the interaction media recorded by an audio or video recording device;
 - determining an at least one pivot spot, being a time position, on the interaction media;
 - determining the time limits of the at least one segment of the interaction media to be analyzed, said limits defining an initial region of interest within the interaction;
 - reducing the initial region of interest by performing an at least one first audio analysis of a first audio analysis type on the initial region of interest and reducing the initial region of interest in accordance with a result of the at least one first audio analysis, to obtain a region of interest; and
 - performing an at least one second audio analysis of a second audio analysis type on the region of interest, wherein the first audio analysis type and the second audio analysis type are selected such that the second audio analysis type requires more computing resources than the first audio analysis type.
20. (Cancelled)
21. (Previously Presented) The method of claim 19 further comprising the step of selecting the first audio analysis or the second audio analysis is based on the at least one event associated with the interaction.
22. (Cancelled)

23. (Previously Presented) The method of claim 19 further comprising the step of selecting a parameters for the first audio analysis or the second audio analysis.
24. (Cancelled)
25. (Previously Presented) The method of claim 19 wherein the region of interest is predetermined by an apparatus.
26. (Original) The method of claim 19 further comprises the steps of receiving interaction data and associated meta-data from an at least one interaction.
27. (Previously Presented) The method of claim 19 wherein the first audio analysis or the second audio analysis comprises analyzing speech elements of the interaction data for the presence of pre-defined words or phrases.
28. (Previously Presented) The method of claim 19 wherein the first audio analysis or the second audio analysis comprises analyzing speech elements of the interaction data to detect positive and negative emotions.
29. (Previously Presented) The method of claim 19 wherein the first audio analysis or the second audio analysis comprises analyzing speech elements of the interaction data for pre-defined speech patterns.
30. (Previously Presented) The method of claim 19 further comprises the steps of
 - identifying an at least one pre-defined computer telephony integration event in the interaction data; and
 - identifying an at least one pre-defined screen event in the interaction data.
31. (Cancelled)
32. (Previously Presented) The method of claim 19 further comprises performing an at least one content analysis step during capturing of the interaction data and the interaction meta-data.
33. (Previously Presented) The method of claim 19 wherein the at least one pivot spot or the region of interest are determined based on an event external to the interaction.

34. (Previously presented) The apparatus of claim 1 wherein the pivot spot is determined using at least one item selected from the group consisting of: a Computer Telephony Integration event; a screen event; an emotional level; and a spotted word.
35. (Previously presented) The method of claim 19 wherein the pivot spot is determined using at least one item selected from the group consisting of: a Computer Telephony Integration event; a screen event; an emotional level; and a spotted word.
36. (Previously Presented) The apparatus of claim 1 wherein the first audio analysis component used for reducing the initial region of interest is selected from the group consisting of: a speaker separation component, emotional level analysis component, word spotting analysis component, audio event analysis component, dual tone mufti frequency (DTMF) event analysis component, and event priority analysis component.
37. (Previously Presented) The method of claim 19 wherein reducing the initial region of interest is done according to an item selected from the group consisting of: speaker separation, audio analysis, emotional level analysis, word spotting analysis, audio event analysis, DTMF event analysis, and event priority analysis.
38. (Previously presented) The apparatus of claim 1 wherein the captured interaction is between an agent and a customer.
39. (Previously presented) The method of claim 19 wherein the interaction media captures an interaction between an agent and a customer.
40. (Cancelled)
41. (Previously presented) The method of claim 19 wherein the method is used for detecting customer churn indications, wherein the pivot spot is defined using a CTI hold event or a cancellation-related screen event; and wherein the region of interest is defined using emotion analysis or word spotting.

42. (Previously presented) The method of claim 19 wherein the method is used for verifying that an agent requested a customer's permission to put the customer on hold, wherein the pivot spot is the time the agent put the customer on hold, the initial region of interest is the whole interaction, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold.
43. (Previously presented) The method of claim 19 wherein the method is used for measuring the effectiveness of a promotion offer to a customer requesting the termination of the service, wherein the pivot spot is the time of a screen event related to offering a promotion or to an account being saved or lost, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot.
- 44-45. (Cancelled).
46. (Previously Presented) The apparatus of claim 1 wherein the at least one pivot spot or the region of interest are determined based on an event external to the interaction.
47. (Previously Presented) The method of claim 19 wherein the reducing step is repeated two or more times.

REMARKS

Claims 1-3, 9-10, 12, 15-17, 19, 21, 23, 25-30, 32-39, 41-43 and 46-47 presented for examination upon entry of the present amendments. Claims 1 and 19 have been amended. Claims 1 and 19 are independent.

Claim rejections, 35 U.S.C. § 103

In the office action, claims 1-3, 9, 10, 15-17, 19, 21, 23, 25-27, 29, 32-39, 41 and 46-47 have been rejected under 35 U.S.C. §103(a) as being unpatentable over U.S. Pre-grant Publication 2004/0083099 by Scarano et al. (hereinafter "Scarano") in view of U.S Patent 6,539,087 to Walsh et al. (hereinafter "Walsh").

Applicants respectfully traverse this rejection. However, Applicants have amended claim 1 to further clarify and make explicit the features of the invention that distinguish over the background art.

Present claim 1 provides an audio or video recording device for recording the audio interaction and obtaining an interaction media; a pivot spot defining component for automatically marking an at least one time position in the audio interaction that indicates the occurrence of an at least one pre-defined event or data item; a first audio analysis component of a first audio analysis type; a region of interest defining component for defining an initial region of interest, by determining the time limits of an at least one segment of the audio interaction, the segment containing the time position of a pivot spot, and for activating the first audio analysis component on the initial region of interest for dynamically reducing the time limits of the initial region of interest to obtain the region of interest; and a second audio analysis component of a second audio analysis type for analyzing the region of interest of the audio interaction. The first audio analysis type and the second audio analysis type are selected such that the second audio analysis type requires more computing resources than the first audio analysis type.

Scarano in view of Walsh does not teach or suggest claim 1. For example:

Neither Scarano nor Walsh discloses that the analysis types are selected such that the second analysis type requires more computing resources than the first audio analysis type. Walsh teaches assigning calls to DSP resources. In other words, the nodes or resources in Walsh are the platforms or machines that perform analysis, and not analysis types that are to be performed as required by claim 1. Thus, Walsh does not relate to multiple analysis types.

Further, Walsh teaches assigning the calls to DSP resources based on the capacities of the resources and not based on analysis type difference. The computing resource requirements of the various analysis types in the disclosed invention are characteristics of the analysis type, as described for example at ¶0021 of the application as published. However, the capacities of the DSP resources in Walsh depend also on their current load and can therefore change over time, *see*, for example, claim 1 of Walsh: "*identifying a first resource having a predetermined capacity to receive **additional** conferences...mapping the channel to one of the plurality of channels of the first resource if the capacity of the first resource is sufficient to **add** the channel[.]*" (Emphasis supplied).

Even further, no difference between the various nodes is taught, *see* Walsh at col. 3 line 25: "[i]t will be appreciated that each of the DSP units 102 of FIG. 1 may include identical or similar circuitry and functionality, although only one of the DSP units 102 is shown in detail." Similarly, no difference between the DSP resources is taught. All parts of all calls in Walsh undergo the same processing no matter which resources they are mapped to, and no selection of different types of analysis are taught, let alone analysis types that are selected such that the second one requires more resources than the first one.

Neither Scarano nor Walsh discloses first and second audio analysis components. Scarano uses a word spotting engine, followed by an SQL query mechanism for retrieving information revealed by the word spotting. Refer, for example, to Scarano at ¶0067: "*The present invention integrates a search for spoken words, phrases or sequences of words in an audio segment with a*

search for traditional SQL data." Walsh also does not teach analyzing the audio by two audio analysis components.

On the "Response to Arguments" section on page 3 of the Office Action, the Examiner asserts that Scarano teaches a first and second audio analysis components and points at elements 1904 and 1905 of Fig. 19. Applicants respectfully disagree. Element 1904 is a decision step, in which it is decided whether to process the audio, based on data external to the audio. *See* Scarano at ¶0159: "[i]f there is no CTI data some information may be provided by the recording device at 1902 such as which phone extension or trunk provided the audio. If the optional CTI interface is included, there is additional data as noted in connection with 1903. Using all available data logic is executed at 1904 and a decision is made about the audio segment." Walsh does not cure this deficiency and does not teach two analysis components of two types.

Neither Scarano nor Walsh discloses that the first audio analysis is activated on a part of the interaction (being the initial region of interest of an interaction). In Scarano, the audio analysis is performed over the full interaction, and not only on a region of interest as required by present claim 1. *See, for example,* ¶0098: "[t]he search set identifies a set of voice communications (e.g., telephone calls) within the speech repository 201. For each voice communication in the set identified by the meta-SQL search, a speech search is executed by the search engine 205 for each of the search expressions that were given in the original search criteria".

On the "Response to Arguments" section on page 3 of the Office Action, the Examiner asserts that Scarano teaches audio segments extracted from a conversation. Applicants respectfully disagree. Scarano does not describe at ¶¶0157, 0159, or other portions, the selection of a part of the conversation. Scarano simply uses the term "segment" to describe the stored audio of the interaction. If Scarano would have meant a segment to relate to a part of the interaction, it would be important and obvious for Scarano to mention the time window within the interaction represented by the segment, which Scarano does not. Further, there is no teaching in Scarano of selecting a part of the interaction that makes it a region in which there is particular interest. Walsh also does not

disclose or suggest processing only part of the data, since it deals with live conferences, in which the system can not decide to drop parts of a conference.

Neither Scarano nor Walsh discloses that the second audio analysis is activated on a part of the interaction (i.e., the region of interest of an interaction). As discussed above, Scarano does not teach a first and second audio analysis types, and also does not teach processing only part of the interaction. Thus, Scarano cannot suggest performing a second audio analysis on a part of the interaction. The second analysis of Scarano, being querying, is performed over data collected on the first phase and not over the audio. Walsh also does not and teach or suggest processing, such as transferring only part of the data.

Neither Scarano nor Walsh discloses activating the first audio analysis component for dynamically reducing the initial region of interest to obtain the region of interest. Scarano uses the audio analysis to obtain words spoken in the interaction, and not to define a region of interest. Walsh does not disclose audio analysis, nor does Walsh disclose operating on part of a conference, so naturally Walsh does not disclose operating an engine to reduce an initial region of interest of the interaction.

In view of the above, the cited combination of Scarano and Walsh does not disclose claim 1, a first and second audio analysis components of a first and a second audio analysis types, activating the first analysis component for dynamically reducing the time limits of the initial region of interest to obtain the region of interest; the second audio analysis component analyzing the region of interest of the audio interaction; or that the second analysis type requires more computing resources than the first analysis type. Reconsideration and withdrawal of the §103 rejection of claim 1 are respectfully requested.

Claims 2-3, 9, 10, 15-17, 34, 36, 38, and 46 depend from claim 1 and, for at least the reason of such dependence, are also patentable over the cited art.

The same arguments as for claim 1 are also applicable to claim 19. Scarano in view of Walsh does not teach or suggest: a first audio analysis and a second audio analysis of a first and second audio analysis types; the first audio analysis being activated on a part of the interaction; the second audio analysis being activated on a part of the interaction; activating the first audio analysis component for dynamically reducing the initial region of interest; and that the second audio analysis type requires more resources than the first audio analysis type. Reconsideration and withdrawal of the §103 rejection of claim 19 are respectfully requested.

Claims 21, 23, 25-27, 29, 32-33, 35, 37, 39, 41, and 47 depend from claim 19 and are also allowable for the reasons set forth with respect to claims 1 and 19 addressed above.

The Office Action rejects claim 30 under §103 as being unpatentable over Scarano in view of Walsh and further in view of U.S. Pat. 6,937,706 to Bscheider et al. ("Bscheider"). In forming the section 103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach the usage of screen events, and introduces Bscheider for this proposition. Assuming *arguendo* that Bscheider so teaches, nevertheless Bscheider does not operate to overcome the several inabilities of Scarano in view of Walsh to disclose independent claim 19, from which claim 30 depends. Thus claim 30 depends from a claim that is allowable, and is, at least by virtue of such dependence, also patentable over the cited art. The dependent claim also contains additional features absent from the prior art of record. For example, claim 30 requires identifying a pre-defined screen event in the interaction data. Screen events relate to events occurring on the screen of the agent, *see*, for example, ¶0019 of the Specification: "[s]creen events are based entirely on what takes place on an agent's display screen. Screen events may be used as triggers to other actions whenever an event of choice takes place. Interactions are tagged with the event, enabling ready search, retrieval and evaluation of the calls. One non-limiting example of a screen event analysis involves the capturing of a field displayed on the agent's screen that indicates the change of status of a user account. For example, when the account status changes from 'Active' to 'Inactive' an event is generated and recorded to a database."

Bscheider, however, relates to screen image data and not to screen events as information additional to the data being processed or as indicators to identifying time locations within the sequence. Applicants are respectfully requesting that the §103 rejection of claim 30 be reconsidered and withdrawn.

The Office Action rejects claims 12 and 28 under §103 as being unpatentable over Scarano in view of Walsh and further in view of U.S. Pub. 2002/0194002 to Petrushin ("Petrushin"). In forming the section 103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach an emotion analysis component, and introduces Petrushin for the proposition that Petrushin teaches an emotion analysis component. Assuming, *arguendo*, that Petrushin so teaches, nevertheless Petrushin does not operate to overcome the several inabilities of Scarano and Walsh to disclose the independent claims. Thus claims 12 and 28 depend from claims that are allowable, and are, at least by virtue of such dependence, also patentable over the cited art. Applicants are respectfully requesting that the section 103 rejection of claims 12 and 28 be reconsidered and withdrawn.

The Office Action rejects claim 42 under §103 as being unpatentable over Scarano in view of Walsh, and further in view of U.S. Pat. 6,724,887 B1 to Eilbacher ("Eilbacher"). In forming the §103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach that the method is used for verifying that an agent requested a customer's permission to put the customer on hold, wherein the pivot spot is the time the agent put the customer on hold, the initial region of interest is the whole interaction, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot and a second predetermined number of seconds following the hold, and introduces Eilbacher for the proposition that Eilbacher so teaches. Assuming, *arguendo*, that Eilbacher indeed so teaches, nevertheless Eilbacher does not operate to overcome the several inabilities of Scarano in view of Walsh to disclose the independent claims. Thus claim 42 depends from a claim that is allowable, and is, at least by virtue of such dependence, also patentable over the cited art.

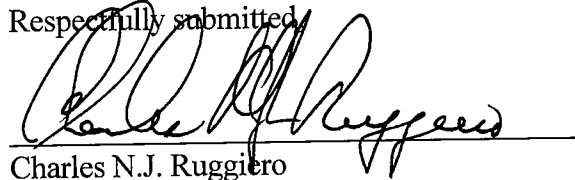
In addition, Eilbacher does not disclose or suggest identifying a particular time location within the interaction. On the contrary, Eilbacher teaches analyzing interactions **as a whole**, see for example Eilbacher at col. 10 line 17: "*these types of recordings allow for evaluating of the full customer experience during the interaction.*" (emphasis supplied). Eilbacher is recording and evaluating full interactions only, cradle-to-grave, and does not teach or suggest setting a pivot spot, which would be meaningless in such a recording scheme. Applicants are respectfully requesting that the §103 rejection of claim 42 be reconsidered and withdrawn.

The Office Action rejects claim 43 under §103 as being unpatentable over Scarano in view of Eilbacher and Walsh, and further in view of U.S. Pat. 5, 918,213 to Bernard et al. ("Bernard"). In forming the §103 rejection, the Office Action acknowledges that Scarano in view of Walsh fails to teach that the method is used for measuring the effectiveness of a promotion offer to a customer requesting the termination of the service, wherein the pivot spot is the time of a screen event related to offering a promotion or to an account being saved or lost, and wherein the region of interest is defined by a first predetermined number of seconds prior to the pivot spot. The Examiner introduces Eilbacher and Bernard to cure this deficiency of Walsh and Scarano. Assuming *arguendo* that Eilbacher and Bernard indeed so teach, nevertheless Eilbacher and Bernard, either separately or in combination do not operate to overcome the inability of Scarano in view of Walsh to disclose the independent claims. Thus claim 43 depends from a claim that is allowable, and is, at least by virtue of such dependence, also patentable over the cited art. Applicants are respectfully requesting that the §103 rejection of claim 43 be reconsidered and withdrawn.

Applicants submit that the claims now pending define patentably over the art of record.
Passage of the claims to allowance is earnestly solicited.

Date July 26, 2010

Respectfully submitted,



Charles N.J. Ruggiero

Reg. No. 28,468

Attorney for the Applicants

Ohlandt, Greeley, Ruggiero & Perle, L.L.P.

One Landmark Square, 10th Floor

Stamford, CT 06901-2682

Tel: 203-327-4500

Fax: 203-327-6401